LUND UNIVERSITY

Gestural viewpoint signals referent accessibility

Debreslioska, Sandra; Özyürek, Asli; Gullberg, Marianne; Perniss, Pamela

Published in:
Discourse Processes

DOI:
10.1080/0163853X.2013.824286

2013

Link to publication

Citation for published version (APA):
Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. (2013). Gestural viewpoint signals referent accessibility. *Discourse Processes*, *50*(7), 431-456. https://doi.org/10.1080/0163853X.2013.824286

Total number of authors:
4

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Gestural viewpoint signals referent accessibility

Sandra Debreslioska[1], Asli Özyürek[2,3], Marianne Gullberg[1], Pamela Perniss[3,4]

[1] Lund University

[2]Radboud University Nijmegen

[3]Max Planck Institute for Psycholinguistics

[4]Deafness, Cognition and Language Research Centre

Word count:

8003

Corresponding author:

Sandra Debreslioska

Postal Address:

Centre for Languages and Literature

Lund University

Helgonabacken 12

223 62 Lund

Phone:

+46 (0)46 222 83 76

Fax:

+46 (0)46 222 32 11

Email:

Sandra.Debreslioska@ling.lu.se

Gestural viewpoint signals referent accessibility

Abstract

The tracking of entities in discourse is known to be a bimodal phenomenon. Speakers achieve cohesion in speech by alternating between full lexical forms, pronouns, zero anaphora as they track referents. They also track referents in co-speech gestures. In this study, we explored how viewpoint is deployed in reference tracking focusing on representations of animate entities in German narrative discourse. We found that gestural viewpoint systematically varies depending on discourse context. Speakers predominantly use character viewpoint in maintained contexts and observer viewpoint in re-introduced contexts. Thus, gestural viewpoint seems to function as a cohesive device in narrative discourse. The findings expand on and provide further evidence for the co-ordination between speech and gesture on the discourse level which is crucial to understanding the tight link between the two modalities.

Acknowledgements

We gratefully acknowledge financial support from the Dutch Science Foundation (NWO) VIDI program and ASPASIA grant awarded to the second author, the Max Planck Institute for Psycholinguistics, and Lund University. We would also like to thank the members of the Gesture and Sign Language lab in Nijmegen, the audience of the Fifth ISGS Conference in Lund for their feedback, and Lian van Hoof for reliability coding.

In order to produce intelligible narratives, speakers at all times have to specify who does what to whom, where and when. Speakers must refer back to entities after their introduction into the discourse using appropriate referential expressions choosing between full lexical forms (e.g. 'the mouse'), pronouns (e.g. 'it') and zero anaphora. However, reference tracking must not only be managed correctly in speech, but it is also reflected in co-speech gestures. A number of studies have shown that speakers' gesture production reflects the tracking of referents in spoken discourse, making reference tracking a bimodal phenomenon (Gullberg, 1998, 2003, 2006; Levy & McNeill, 1993; Marslen-Wilson, Levy & Komisarjevsky Tyler, 1982; McNeill & Levy, 1992;). These studies have demonstrated that speakers produce co-speech gestures when introducing and re-introducing entities with full lexical expressions, whereas they tend not to produce gestures when maintaining referents from one utterance to the next with pronouns or zero anaphora. Speakers thus create an opposition between referents with different information status in both modalities. Furthermore, speakers might choose to use gestures as a cohesive device by creating visual co-reference in space. Each animate referent can be associated with a particular locus in gesture space which the speaker can use consistently when the entity is being gestured about throughout the discourse. Speakers basically charge concrete gesture space with referential meaning and turn it into a map of discourse (Gullberg, 2006).

What remains unclear is whether and if so how speakers also represent referents differently in gesture as a function of reference tracking. Specifically, this study will look at how viewpoint is deployed in reference tracking focusing on representations of animate entities and their actions. The starting point is the consideration that speech and gesture are two aspects of the utterance that are produced together and as such form an audiovisual 'ensemble' (Kendon, 2004), a 'composite signal' (Clark, 1996) or an 'integrated system' (McNeill, 1992) within which the two modalities are closely linked in a temporal, semantic/pragmatic, syntactic and discursive way. This study will explore how the representational means of gestures are affected by discourse organizational principles, and more specifically, we examine whether and how co-

speech gestures show sensitivity to the discourse accessibility of the animate referents they represent. In the following, we outline how accessibility can be measured in spoken contexts, we then provide some background on gesture and its relation to speech, and finally we present the current study.

Reference tracking in speech

After introducing entities into discourse for the first time, speakers need to refer back to them in an appropriate way in order for listeners to understand what is referred to. Languages offer speakers various linguistic means to indicate the information status of discourse entities. To refer to an animate referent, speakers can, for instance, use nouns, pronouns or zero anaphora. The mechanism that leads us to prefer a specific referential device over another at a particular moment in discourse lies in the accessibility of the entity referred to – its cognitive status or givenness – which is supported by a large body of research for various languages (e.g. Ahrenholz, 2005, 2007; Ariel, 1988, 1990, 1994, 1996, 2001; Chafe, 1994; Givón, 1983, 1984; Gundel, Hedberg & Zacharaski, 1993; Hendriks, 2003; Hickmann & Hendriks, 1999; Lambrecht, 1994; for an overview, see e.g. Huang, 2000). This set of studies argues for a more or less universal principle which sees the interaction between the linguistic means used to refer and the accessibility of the entity as central. In this view, entities that represent 'new' information to the listener and are therefore assumed to be less accessible will be marked with fuller forms. In contrast, entities that represent information that has already been introduced, i.e. 'given' information, and are therefore more accessible to the listener, will be marked by a less full form. Givón (1984) posits a universal principle of quantity which states that "more continuous, predictable, non-disruptive topics will be marked by less marking material; while less continuous, unpredictable/surprising, or disruptive topics will be marked by more marking material" (Givón, 1984: 126). According to this principle, zero anaphora show the strongest relationship to continuous topics (known, given, more accessible information), while full lexical noun phrases (NPs) are most likely related to discontinuous topics (less known, newer, less accessible information).

An operationalization of the accessibility of referents is local co-reference which is defined by whether a protagonist is referred to in two consecutive utterances or not (Hickmann & Hendriks, 1999). If this is the case, the referent is considered to be highly accessible; it is maintained from one utterance to the next. If this is not the case, the referent is still accessible but to a lesser degree and must be re-introduced. This variation in accessibility is then reflected in the linguistic forms (i.e. referential expressions) used to refer to them. Less full linguistic forms, such as zero anaphora, are used when the referent has been mentioned in the immediately preceding utterance (i.e. the referent is immediately maintained) (example 1), and fuller forms, such as full lexical NPs, are used when the referent has not been mentioned in the preceding utterance (i.e. the referent is re-introduced) (example 2).

(1)

1 Die Maus$_1$ macht Rollen um das Reck

2 Und ø$_1$ springt dann ab.

'1 The mouse$_1$ rolls on the bar

2 And then ø$_1$ jumps off.'

(2)

1 Die Maus$_1$ macht Rollen um das Reck.

2 Der kleine Elefant$_2$ schaut zu

3 Und dann springt die Maus$_1$ ab

'1 The mouse$_1$ rolls on the bar.

2 The little elephant$_2$ observes that

3 And then the mouse$_1$ jumps off.'

Gestures and reference tracking

In the field of gesture studies, gestures (also labelled gesticulations or co-speech gestures) are characterized as expressive and communicatively relevant movements of the

hands and arms that are related to the speaker's ongoing talk. Gestures and speech are semiotically and symbolically linked in many ways. They create and express meaning together (e.g., semantically and pragmatically) and in close temporal proximity.They are therefore considered to be tightly linked (cf. Kendon, 2004; McNeill, 1992).

The close relationship between speech and gesture has further been shown to be reflected in, for instance, the parallel development of speech and gesture in childhood (e.g. Jancovic, Devoe & Wiener, 1975), the parallel breakdown in dysfluency (Seyfeddinipur, 2006) and in the variation that co-speech gestures show with different typological and structural patterning of spoken languages (Kita & Özyürek, 2003). It crucially also extends to discourse.

A set of studies has provided evidence that reference tracking in narratives is characterized by systematic and repeated associations between specific gestural behavior and referential expressions (e.g. Levy & McNeill, 1993; Marslen-Wilson et al. 1982). Gestures have been shown to reflect Givón's (1983, 1985) principle of quantity because they are said to increase in number when the linguistic quantity of nominal referential forms they accompany become fuller (McNeill & Levy, 1992). A range of studies has shown that when characters are first introduced into discourse (and are new, less continuous and accessible) they tend to be expressed by a lexical NP (cf. also e.g. Chafe 1994; Givón 1983) and to be accompanied by a gesture (Gullberg, 1998, 2003, 2006; Levy & Fowler, 2000; Levy & McNeill, 1993; Marslen-Wilson et al. 1982; McNeill & Levy, 1992). In contrast, when referents are maintained (and are given, more continuous and accessible) speakers are inclined to use the most reduced form that is sufficient to specify the intended referent to the listener, that is, pronouns or zero anaphora. These forms in turn are usually not accompanied by any gesture. However, speakers sometimes choose to refer to a maintained entity with a fuller referential form. The use of a lexical NP for a subsequent mention of a referent can serve multiple functions and be interpreted in different ways. One explanation, as seen above, is that the entity referred to is re-introduced, meaning that an intervening subject and/or a certain distance between the present and last mention of the referent has occurred (cf. Hickmann & Hendriks, 1999; but see also Clancy, 1980 and

Gullberg, 2003). Re-introduced referents are also more likely to coincide with the use of a gesture.

Furthermore, gestures co-occurring with referential expressions have been shown to work together with space to create cohesion by providing spatial anaphoric linkages in narratives (e.g. Gullberg, 2003, 2006). This can be called spatial anaphoricity. When referents are first introduced into discourse, speakers may accompany the NPs with gestures that are anchored or localized in a part of space. When a given referent is next mentioned, speakers often gesture towards the same locus in space, previously established as associated with the referent. Speakers thereby visually re-activate the location and thus the referent associated with it. These gestural re-activations predominantly co-occur with full NPs (i.e. the re-introduction of referents; Gullberg 2003, 2006). Other studies (So, Kita & Goldin-Meadow, 2009) have also indicated that gestures that represent referents are consistently located in parts of space previously associated with them. However, these studies have not regarded the referents' degree of accessibility in discourse terms (i.e. whether they are maintained, re-introduced, etc.).

Kommentar: this now feels a little disconntected here. As a reader, I'm wondering 'why are you telling me this now?'

Overall, then, this line of research has suggested that gestures are used to create cohesive linkages in connected discourse, whereby speakers can vary the number of gestures but can also consistently use them in space in order to indicate differences in information status of referents, maintained or re-introduced (cf. also Levy & Fowler 2000).

What remains unclear, however, is whether – and if so how – gestures signal the accessibility of referents in they way they represent those referents. Representational gestures have been defined as gestures that represent entities by using indexical or iconic means (e.g. McNeill, 1992; Kita, 2000). Representational gestures that are indexical in nature are also called deictic or pointing gestures. They can be concrete, indicating objects present in the environment, or abstract, pointing to a seemingly empty space , as if the speaker is establishing a virtual object/entity or pointing (back) at such a virtual object/entity. Representational gestures that are iconic in nature on the other hand, show a certain degree of isomorphism between the entities that they represent (Kita, 2000), usually depicting their properties (e.g. shape, size),

their relationship to other entities (e.g. next to, above), and/or their actions (e.g. rolling, falling). Furthermore, representational gestures of iconic nature are usually produced from a certain viewpoint. Gesture studies have traditionally discussed character viewpoint (CVPT) and observer viewpoint (OVPT) (e.g. McNeill, 1992; Parill, 2009). CVPT gestures reflect mappings of the protagonist's body onto the body of the speaker. When producing these gestures, the speaker is enacting the protagonist and/or what she is doing by using her own body. The speaker seems to assume an insider's perspective as her body becomes part of the gesture space. OVPT gestures, on the other hand, do not contain any enactments. Rather, the speaker seems to be looking onto the scene from the outside in contrast to being part of it. The speaker uses her hand(s) to represent the protagonist as a whole. For instance, two hands can be used to mould the shape of a protagonist or a finger can be used to draw a path that a protagonist is moving along.

Some suggestions in the literature point towards a possible co-ordination between gestural viewpointand discourse. Firstly, McNeill (1992) and Parrill (2009) have suggested that gestural viewpoint is connected to the transitivity of co-occurring verbs. CVPT viewpoint gestures have been found to be related to transitive verbs and OVPT gestures to intransitive verbs. Parrill (2012) has also recently shown that in a context where speakers and listeners share knowledge about an event (i.e. when they have watched a stimulus together), speakers use fewer CVPT gestures compared to contexts where they do not share knowledge. Furthermore, McNeill (1992) has proposed a scale of gestural modes of representation progressing from no gesture, via beats (i.e., rhythmic movements used to emphasize a certain segment in the discourse), pointing gestures (i.e., points to real objects in the speaker's environment, or abstract entities in gesture space), and different OVPT gestures, to CVPT gestures (beats – pointing – OVPT – CVPT). Importantly, he relates his scale to Givón's (1985) principle of quantity and suggests that the more linguistic material is used to encode a referent – as in going from zero anaphora via pronouns, stressed pronouns, different kinds of full NPs to predicates – the higher the speaker will move up on the gesture scale. This in turn means that OVPT gestures are

more likely to align with different kinds of full NPs, and CVPT gestures with verbal elements of the clause.

However, none of the previous studies have looked specifically or quantitatively at the relationship between viewpoint and the status of the gestured referents in discourse (i.e., in maintained vs. re-introduced contexts). This study therefore focused on gesture viewpoint in order to examine whether speakers use this level of representation in a systematic way to mark the difference in referents' accessibility over discourse. In addition, contrary to previous studies, we examined gestures across referential contexts, defined as clauses, rather than examining gestures in exact temporal alignment with referential expressions in a narrow sense (as in e.g. Gullberg, 1998, 2003, 2006; Marslen-Wilson et al. 1982).

The current study

The purpose of this study is to investigate how referents with different degrees of accessibility are represented in gesture. More specifically, the study will focus on viewpoint to examine whether speakers use this level of representation in a systematic way to mark the difference in information status of a discourse referent (maintained versus re-introduced). We asked the following question: Do speakers mark the difference between referents in maintained contexts versus referents in re-introduced contexts in gesture viewpoint?

For speech, we know that a formal distinction is made between referents in the two contexts. Therefore, we expected to replicate robust findings to the effect that maintained referents will mostly be encoded with reduced referential expressions, such as pronouns and zero anaphora, and that re-introduced referents will mostly be encoded by full lexical NPs.

For gesture, we derived the following predictions from McNeill (1992). McNeill relates different kinds of OVPT gestures to different kinds of full NPs and relates CVPT gestures to predicates or verbal elements of the clause. If we consider that referents in a highly accessible

(i.e. maintained) context are usually encoded with reduced referential expressions, such as zero anaphora, we can infer that the main linguistic element that is referring to the entities in question – in other words, the main spoken forms that say something new about the referents – will be the verbal elements of the clause (Note that the verbal element is not a referential expression per se, but is the only linguistic element that expresses referent relevant information). In contrast, in less accessible contexts, referents are more likely to be encoded with fuller referential expressions, such as full NPs. Therefore, we propose the following predictions as inspired by McNeill's (1992) gesture scale. We should find more CVPT gestures in maintained (i.e. more accessible) contexts because the referents in this context are very likely to be encoded with a zero anaphor in speech. It follows that the main referential form relating to the referent will be the predicate and thus, there should be an increased likelihood of a CVPT gesture occurring in this context. In contrast, when speakers re-introduce referents, i.e. referents which are less accessible, they typically encode them with lexical NPs. In re-introduced contexts, therefore, we should find OVPT gestures to be more likely.

Viewpoint has also been shown to be connected to the transitivity of the verbs/clauses that the gestures accompany (McNeill, 1992; Parrill, 2010). As such, CVPT gestures have been found to predominantly accompany transitive and OVPT gestures to accompany intransitive verbs. We will also take this relationship into account, and test whether the effect of transitivity has an independent effect on gestural viewpoint.

Thus, the predictions for speech and gestures related to reference tracking of animate entities can be summarized as follows.

First, we expect to find a similar pattern as in previous studies on the spoken forms for reference tracking of animate referents in discourse. Specifically, referents in a maintained context will mostly be encoded with reduced referential expressions, such as pronouns and zero anaphora. Referents in a re-introduced context will be more likely to be encoded by full NPs. Secondly, we expect speakers to use CVPT gestures to represent referents in a maintained context, and OVPT gestures to represent referents in a re-introduced context. Finally, we expect speakers to use

CVPT gestures when describing transitive events and OVPT gestures when describing intransitive events. We also hypothesize that the effect of transitivity be independent from the effect of referential context.

METHOD

Participants

Ten students enrolled at the European-University Viadrina in Frankfurt (Oder), Germany participated in the study (3 males, median age 24; 6). All participants were native speakers of German. They were recruited at the University through notices, or by way of mouth through personal networks. Each participant was paid 8 Euros per hour.

Procedure

Participants were invited in pairs into a quiet room at university. They were randomly assigned the role of speaker or listener. The experimenter gave them instructions orally. The task for the speaker was to watch different cartoons presented on a laptop with sound turned off. After each clip, the speaker had to turn away from the laptop (which displayed a blank white screen between clips), and face forward toward the addressee to narrate the episode just seen. Speakers were allowed to watch cartoons more than once in case they needed more time to memorize the story. During their narration the addressee was not allowed to interrupt. After each narration, addressees were asked to retell the story based on the single production given by the speaker. Three cameras were used to video tape speakers' production for later analysis. Two cameras captured front views of the speaker and addressee, and a third camera (fitted with a wide-angle lens) was positioned above the two participants in order to capture a top view of both.

Stimulus materials

Seven short films were used to elicit narratives. Five of the films were taken from a German children's television series called "Die Sendung mit der Maus" (Westdeutscher

Rundfunk, 1972; also used in Perniss, 2007 for German sign language elicitation), and feature a large personified mouse and a smaller personified elephant as the main protagonists. In three of the stories, only the mouse and the elephant are present. The fourth story also features a man in a green suit and the fifth only features the elephant together with a smaller duck and a larger hippo. The other two films were short silent stories starring Charlie Chaplin (both from "City Lights", 1931; also used in So et. al, 2009). One of them features Charlie Chaplin and a woman as the main protagonists. An additional animate character, a cat, also plays a role in this film. In the second story, Charlie Chaplin and another man are the only characters (see appendix for descriptions of the different films). Each participant retold all 7 films, yielding 70 narrations in total for analysis from the 10 participants.

Speech annotation and coding

All 70 narratives were transcribed using German standard orthography and were divided into clauses. A clause was defined following Bermann & Slobin (1994) as "any unit that contains a unified predicate … a predicate that expresses a single situation (activity, event, state). "(p. 660). Embedded clauses, as in (3), were treated as separate clauses, meaning that example (3) contains two clauses. This procedure was relevant for the coding of gestures as well. Gestures that appeared within the limits of the embedded clause were counted as belonging to this clause whereas the gestures co-occurring with parts of the main clause were counted as belonging to the latter.

(3)

1 Und dann macht der Elefant, in dem er sich auf die Brust klopft, weiter Musik

'1 And then the elephant continues, by hitting his chest, making music'

Note that we restricted our analyses to reference and representation of animate entities in subject position, meaning that we focused only on those clauses in the narratives that

contained a reference to the animate entities in syntactic subject position. This left us with 539 clauses which were further coded.

Referential form and referential context. Each clause was further coded for:

a) Identification of animate entity referred to in syntactic subject position (e.g. mouse, elephant, Charlie Chaplin, etc.)

b) Identification of its linguistic referential form (e.g. lexical NP, pronoun or zero anaphor)

c) Determination of information status of each referent in syntactic subject position (e.g. (immediately) maintained or re-introduced).

To determine the referential context in which an animate entity was mentioned, we focused on local co-reference across clauses and used the framework by Hickmann & Hendriks (1999) (see figure 1). A maintained context corresponded to cases where a referent was mentioned in two consecutive utterances in subject position (i.e. a coreferential context), and a re-introduced context corresponded to cases where the current mention of an animate entity in subject position was preceded by an intervening animate entity in subject position (non-coreferential context). Note that all introductions/first mentions were excluded from our analysis.

Insert Figure 1 here

A clause was labeled M if the referent it contained was (immediately) maintained, as in examples (4-5), and as RI if the referent it contained was re-introduced, as shown in example (6). Subscripts are used to indicate co-reference.

(4)

1 Und dann kommt von der linken Seite ein Mann$_1$ mit einem Zylinder rein

2 Und er$_1$ geht unter dem Reck durch (M)

'1 And then from the left side, a man$_1$ with a top-hat comes in

2 And he$_1$ passes underneath the bar' (M)

(5)

1 Und die Maus$_1$ versucht Eierkuchen zu machen

2 Ø$_1$ Nimmt die Pfanne weg (M)

3 Ø$_1$ Dreht sich um (M)

'1 And the mouse$_1$ tries to make pancakes

2 Ø$_1$ Takes the pan (M)

3 Ø$_1$ Turns around (M)

(6)

1 Also diesmal kommt die Maus$_1$ von der rechten Seite rein ins Bild

2 Der Elefant$_2$ ist schon in der Mitte

3 Und die Maus$_1$ hat einen roten Ball mit (RI)

4 Und zuerst werfen sie$_{1,2}$ sich den Ball gegenseitig zu (RI)

'1 So, this time the mouse$_1$ comes into the picture from the right side

2 The elephant$_2$ is already in the middle

3 And the mouse$_1$ has got a red ball (RI)

4 And first they$_{1,2}$ throw the ball back and forth to each other' (RI)

With plural pronouns, such as the pronoun *sie* 'they', referring to two referents, the information status was considered to be re-introduced and the clause containing it coded as RI if at least one of the referents was not in subject position in the preceding clause (example 6, 7). Similarly, the information status of a singular form contained in a clause which followed a clause with a plural form in subject position was also considered as re-introduced (example 7).

(7)

1 Der Mann$_1$ begleitet die Frau zu einer Treppe (RI)

2 Und Ø$_1$ überreicht ihr einen Korb (M)

3 Und sie$_{1,2}$ wollen sich verabschieden (RI)

4 Sie$_2$ möchte einfach gehen (RI)

'1 The man$_1$ accompanies the woman to the steps (RI)

2 And Ø$_1$ hands her a basket (M)

3 And they$_{1,2}$ want to say good-bye (RI)

4 She$_2$ just wants to leave' (RI)

Exceptions for the syntactic coding of referents in subject position were made when clauses contained verbs that were used with the subject pronoun *es*, as in *stören* 'disturb', *gefallen* 'like', or *gelingen* 'succeed'. In these instances the animate entity can only be used in object position (corresponds to impersonal sentences constructed with 'it' in English) (see example 8). Thus, clauses that contained verbs used with the subject pronoun *es* were also counted and categorized according to the schema described above.

(8)

[es] gefällt natürlich der Maus nicht

'[it] does not please the mouse of course'

Note that we focused our analyses on clauses on the narrative level (see e.g. McNeill, 1992, for the distinction between different narrative levels and gesture). However, we included specific meta-narrative statements when they contained narrative phrases (cf. Levy & Fowler, 2000), such as in example (9), where 'someone coming down the stairs' is the narrative phrase.

(9)

und da siehste jemanden die Treppe runterkommen

'and you see someone coming down the stairs'

Transitivity of the event. For every clause, we determined the transitivity of the event it referred to in the film. We chose to examine the transitivity of events rather than the transitivity of verbs in the linguistic sense on the grounds of cross-linguistic comparability. Verb transitivity differs crosslinguistically. For example, 'to thank' in English is a transitive verb but *danken* 'to thank' in German is an intransitive verb. To avoid problems of crosslinguistic comparability, we therefore chose to examine the event construction instead. An event was considered to be transitive when it involved a manual/pedal activity of animate entities acting on another entity, inanimate or animate. An event was considered to be intransitive when it involved motion/location of animate entities, by not their acting on any other entity.

Gesture annotation and coding

For the gesture annotation and coding, we used frame-by-frame analysis of digital video in a video annotation software developed at the Max Planck Institute for Psycholinguistics (ELAN, www.lat-mpi.eu/tools/elan/, Sloetjes & Wittenburg, 2008)

All gesture strokes for relevant clauses were annotated. A stroke is defined as the most effortful part of the gesture. It is the peak of the gesture that carries the meaning which the gesture is to express. Put differently, it is the meaningful or expressive part of the gestural movement where the spatial excursion of the limb reaches its apex (cf. McNeill, 1992; Kendon, 2004). To determine onset and offset of gesture strokes, we considered changes in the parameters of shape, placement of the hand, trajectory of motion and tension of the hands (for more on gesture phases and how to recognize and code them, cf. e.g. Kendon, 2004; Seyfeddinipur, 2006).

Representational gestures and meaning. All representational gestures of an iconic nature (i.e., viewpoint gestures) in the relevant clauses were determined. They represented size, shape, actions or movements of the entities. Each gesture was then assigned a meaning indicating which protagonist was represented. The semantic meaning was determined by looking at the form of the gesture form together with the speech in the clause with which it occurred. Specifically, only gestures that could be said to overlap in meaning with the information about the animate referent in subject position in speech on the clause level were considered. An example would be a speaker talking about the mouse circling on a high bar and then dismounting it while first drawing circles in the air and then tracing the path of the mouse dismounting from the bar and landing on the floor with her finger. In difficult cases, gesture form was compared to stimulus film context in order to determine the meaning of the gesture. Gestures that represented referents mentioned in object position, or any inanimate entity, were excluded. This would be the case if a speaker gestured about the "bar" when describing that the mouse was circling on it. Note that sometimes gestures that mainly represented the animate entity in subject position could also include some aspect of another animate or inanimate entity. For instance, a speaker could enact how someone was throwing a ball by also depicting the shape of the ball with the two hands. Rhythmic and emblematic gestures were excluded from further analysis since they do not represent entities or their actions, and deictic or pointing

gestures were excluded since they cannot have a viewpoint (see Kendon, 2004, for an overview of gesture taxonomies and classification schemes).

Viewpoint coding. All relevant gestures were divided into OVPT and CVPT gestures (McNeill, 1992). In CVPT gestures the protagonist's body parts are mapped onto the speaker's body (i.e. hands, body, face/ head). The speaker represents the protagonist by using her own body to depict the character in question. Thus, the gesture incorporates the speaker's body into gesture space. OVPT gestures are cases where the gesture excludes the speaker's body from the gesture space and her hands play the part of the character as a whole (see figure 5 for examples of each viewpoint).

If the speaker mapped some part of the character's body onto her own body, the gesture was coded as a CVPT gesture. If this was not the case, the gesture was coded as an OVPT gesture. Importantly, this procedure eliminates any potential concern about mixed or dual viewpoints (e.g. Parrill, 2009). Our CVPT category comprises every gesture that exhibits a mapping onto the speaker's body. However, note that we treated gestures with dual elements (i.e. those which represented two referents, both of them from a character perspective) as follows. Usually, for these gestures, the speaker's body represents one of the characters and the hands represent another character (cf. Parrill, 2009). We coded only the articulator which corresponded to the referent referred to in subject position as a CVPT gesture.

OVPT gestures mainly consist of gestures depicting the path, manner of motion, shape or size of entities.

Inter-coder reliability

Fifteen percent of the gestures were coded for viewpoint by an independent coder to test inter-observer reliability. The Kappa statistic was performed (epi.kappa in R) to determine consistency among raters. The proportion of agreements after chance had been excluded was 0.87 (95% CI 0.77 to 0.98). There was thus substantial agreement between the two coders. Also,

McNemar's chi-squared test statistic was 1.8 (P = 0.18), meaning that there was little evidence that the two coders determined viewpoint in gestures differently.

Analyses

In total, we had 70 narrations by 10 speakers which contained of 539 applicable clauses. For the speech analysis, we examined the use of full NPs, pronouns and zero anaphora in both contexts (i.e. maintained and re-introduced) in those clauses. The gesture analysis targeted the same 539 clauses as the speech analysis in which 545 applicable gestural strokes occurred. These were representational gestures produced from either a CVPT or OVPT and co-occurred with either a transitive or intransitive event.

We analyzed the data using a multilevel logistic regression model (cf. Baayen, 2008) to predict the likelihood of speakers labelling maintained referents versus re-introduced referents by either full NPs, pronouns or zero anaphora. For gesture, we tested the likelihood of speakers using either CVPT or OVPT gestures to represent referents in maintained versus re-introduced contexts; and finally, the likelihood of speakers using either CVPT or OVPT gestures in transitive versus intransitive events. Note that the graphs will show numbers that are based on proportions. The analyses themselves are performed on raw numbers.

Multilevel modeling is particularly suitable for within-participant designs where large sources of variance across individual participants and/or items and missing data points are expected. This is the case in the current study. People's gesture rates vary considerably between and within participants depending on what is being talked about. Mixed effect logistic regression allows us to account for this variation by including these sources of variance as random effects. Maximum likelihood estimation is used to estimate parameters for each dependent measure for each participant, and this is under the assumption that all participants come from a normally distributed population in that regard (e.g. here for the production of different viewpoint gestures). It then allows us to explore how different factors (i.e. what we will call fixed effects) affect performance above and beyond each participant's individual parameter for the relevant

measures. Similar parameters can be simultaneously estimated for random factors (such as participant and stimulus) even when data is missing, incomplete, or unbalanced (cf. Cook, Yip & Goldin-Meadow, 2012).

RESULTS

Speech results

The speech analyses focused on which referential expressions were used to refer to animate entities depending on their information status as maintained or re-introduced (clauses labeled M and RI, respectively). Figure 2 summarizes the distribution of full lexical NPs, pronouns and zero anaphora in M and RI clauses.

In our multilevel logistic regression model, we included referential context as a fixed effect, and participant and stimulus as random effects and did three pairwise analyses (comparing the use of NPs to pronouns, NPs to zero anaphora and pronouns to zero anaphora in the two contexts, maintained versus re-introduced). Pairwise analyses were needed because multilevel logistic regression models only allow for binary variables (i.e. both independent variables or fixed effects and dependent variables). The analysis examining the use of NPs in comparison to zero anaphora revealed that NPs were used significantly more in re-introduced than maintained contexts compared to zero anaphora (β= 7.4564, SE= 0.8596, $p < 0.000$). Furthermore, the analysis investigating the use of NPs in comparison to pronouns revealed that NPs were also used significantly more in re-introduced than maintained contexts compared to pronouns (β= -2.270, SE= 0.407, $p < 0.000$). Knowing that fuller referential expressions (NPs) were used significantly more in re-introduced contexts and leaner referential expressions (zero anaphora and pronouns) were used significantly more in maintained contexts, we compared the use of zero anaphora to pronouns in order to determine which of these forms was preferred in maintained contexts. The analysis revealed that zero anaphora were used significantly more in maintained than in re-introduced contexts compared to pronouns (β= 4.2566, SE= 0.5566, $p < 0.000$). This means that pronouns were used significantly more in re-introduced contexts as

compared to zero anaphora, but were also used significantly less in re-introduced contexts as compared to NPs.

Thus, the results show that maintained referents were predominantly labelled by zero anaphora, whereas re-introduced referents were mostly labelled by full NPs.

Insert Figure 2

Gesture results

The gesture analyses focused on which gestural viewpoint, CVPT versus OVPT, was used to represent animate entities depending on the contexts they were mentioned in, maintained or re-introduced (clauses labeled M and RI, respectively). Furthermore, we examined which gestural viewpoint, CVPT versus OVPT, was used depending on what type of event was encoded in the relevant clauses, transitive versus intransitive (labeled T and IT, respectively). Figure 3 and figure 4 respectively summarize the distribution of CVPT and OVPT gestures in M and RI clauses and the distribution of CVPT and OVPT gestures in transitive and intransitive events.

In our multilevel logistic regression model, we included referential context and transitivity of events as fixed effects, and participant and stimulus as random effects. The analysis showed that both fixed effects have a significant effect on the use of gestural viewpoint (i.e. the two effects are independent from each other). Concerning gestural viewpoint in relation to referential context, the analysis revealed that CVPT gestures were used significantly more to represent referents in maintained contexts than in re-introduced contexts as compared to OVPT gestures (β= -0.4738, SE= 0.2403, p <.05). Concerning gestural viewpoint in relation to transitivity of events, the analysis revealed that CVPT gestures were used significantly more in transitive events than in intransitive events as compared to OVPT gestures (β= 2.1485, SE= 0.2748, p < 0.000).

Thus, the results show that referents in maintained contexts are predominantly represented by CVPT gestures, whereas referents in re-introduced contexts are mostly

represented by OVPT gestures. Furthermore, the results show that transitive events are predominantly accompanied by CVPT gestures, whereas intransitive events are mostly accompanied by OVPT gestures.

Insert Figure 3

Insert Figure 4

Example

Figure 5 illustrates the phenomenon at hand. The participant has just described how the mouse was rolling on a bar and the elephant came onto the scene, saw what the mouse was doing and was very happy (example 10, lines 1-2). The speaker then explains how the elephant jumped back and forth (example 10, line 3) using a zero anaphor in speech because he is maintaining the elephant from one utterance to the next. In gesture, he produces a CVPT gesture, mapping the elephant's legs onto his arms and enacts the jumping action by moving his arms up and down (Figure 5, picture1 and example 10, line 3; the gesture stroke aligns with the letters marked in bold face). The speaker then shifts to talk about the mouse (example 10, line 4) using a lexical NP because he is re-introducing it. In gesture, he switches to an OVPT representation. He draws the path of the mouse's movement with his finger (Figure 5, picture 2 and example 10, line 4; the gesture stroke aligns with the letters marked in bold).

Insert Figure 5

(10)

1 $Ø_2$ Sieht das (ja)

2 $Ø_2$ freut sich unheimlich

*3 $Ø_2$ springt vo**n Vorder- auf die Rückbeine** und wieder zurück (See figure 5, picture 1)*

4 *und die Maus$_1$ **springt dann i**rgendwann ab (See figure 5, picture 2)*

'1 $Ø_2$ sees that (yes)

2 $Ø_2$ is really happy

3 $Ø_2$ jumps fro**m its forelegs to its back legs** and back again (See figure 5, picture 1)

4 and the mouse$_1$eventually **then jumps** off' (See figure 5, picture 2)

DISCUSSION

This study set out to examine whether speakers systematically use gestural viewpoint to mark the difference between referents in maintained and re-introduced contexts in discourse thereby reflecting discourse organization properties in gesture in this dimension.

For speech, the results showed the expected patterns. That is, speakers used fuller linguistic referential forms when referents were re-introduced and leaner referential forms when referents were maintained. Thus, the results are in accordance with previous research on this topic suggesting that, for reference tracking, German speakers vary the referential expressions in relation to the accessibility of referents (e.g. Ariel 2001). More specifically, German speakers prefer to use full NPs for re-introduced referents and zero anaphora for maintained referents.

For gesture, we found that speakers use viewpoint in a systematic way to signal the accessibility of the referents they are representing. Speakers prefer CVPT gestures for representing referents in maintained/more accessible contexts and OVPT gestures for referents in re-introduced/less accessible contexts. Crucially, these results confirm the predictions derived from McNeill's scale of gestures and show that the accessibility of referents not only influences the choice of referential expressions in speech, but also the representational form of gestures. Note that McNeill (1992) considered the exact temporal alignment between the two modalities but that we have derived predictions for the co-expressivity of speech and gesture at the clause level. The results are also generally in accordance with previous studies about

gestural reference tracking which have proposed that the use of more or less gesture and the anaphoric use of space are used to mark the difference between maintained versus re-introduced referents (e.g. Gullberg, 2003, 2006; Levy & McNeill, 1992; McNeill & Levy, 1993). Similarly, the present study proposes that the difference in referential contexts is marked by viewpoint, suggesting that gestures are sensitive to the change of information status of a referent over discourse and that gestures can come to visually mark these shifts in a systematic way.

Furthermore, we have found that there is an independent effect of event transitivity on the use of viewpoint. These results support previous studies on the relationship between transitivity and viewpoint in gestures (McNeill, 1992; Parrill, 2009) even if we used a different approach, examining the transitivity of events rather than transitivity of verbs.

To summarize, we have replicated findings for speech, showing that speakers signal the accessibility of referents by varying the form of referential expression. More importantly, we have provided the first quantitative evidence that a parallel preference is found in gesture, such that speakers signal the accessibility of referents by varying the use of viewpoint in representational gestures- CVPT for referents in maintained and OVPT for referents in reintroduced contexts.

An obvious question is why speakers should produce more CVPT gestures in maintained and more OVPT gestures in re-introduced contexts. We propose that in maintained contexts, entities to be represented are easily recoverable and do not need to be differentiated from others. This is why in speech speakers use zero anaphora to refer to them. Similarly in gesture, we might expect a representational form that does not incorporate a ready means for referent differentiation. Because CVPT gestures rely on mapping a referent on the speaker's body, and any speaker only has one body on which to accomplish such a mapping, CVPT gestures are good candidates for such a form that does not allow for entities to be differentiated. In CVPT gestures, the speaker's body– in the same location and in the same appearance– has to be used for every entity. In contrast, OVPT gestures allow for more differentiation by making use of the availability

of spatial mapping of referents. The use of space provides a ready means for referent differentiation, more similar to the use of fuller, more specific referential forms in speech. Previous studies have shown how speakers use space to anchor referents at different loci and take advantage of the option to refer back to those referents by reusing the loci previously assigned to them (i.e. by gesturing back towards those loci– an affordance of the visual modality that sign languages also use for the tracking of referents in discourse[1]). Reactivating the location thereby serves to reactivate the referent that was assigned to it. This is to say, OVPT gestures give speakers the possibility to signal the relationships between different entities spatially, track their respective movements, and essentially create visual maps of discourse that can help the listener to more concretely understand who is doing what to whom and where (cf. e.g. Gullberg, 2006; So et al., 2009). When the entity needs to be specified more overtly, such as in re-introduced contexts where speakers mostly use full NPs in speech, they also prefer OVPT gestures.

Finally, note that viewpoint cannot always be divided in a dichotomous way. Dual or mixed viewpoints have characteristics from both kinds of viewpoint. An example of a mixed viewpoint is when the hand is representing the trajectory of an inanimate entity, such as a ball that is flying towards a person, while the body of the speaker represents the person who is being approached by the ball. Mixed viewpoint might be considered to be a separate dimension. It remains to be explored whether mixed viewpoint gestures might also function differently from CVPT or OVPT gestures in the representation of entities with different degrees of accessibility.

Another important aspect to be investigated further is the issue of temporal alignment between gestures and speech. McNeill's (1992) scale of gestures relied on exact temporal alignment between gestures and referential expressions. It suggests that CVPT gestures should mostly co-occur with predicates whereas OVPT gestures should rather align with NPs. In the current data we have observed exceptions to this proposed tendency. This may be because we

[1] But note that some sign languages, such as German Sign Language (DGS), have devised systematic means for differentiating referents in CVPT (e.g. by incorporating a meaningful turn of the shoulders into the form).

have focused on representational gestures – mostly movement and action gestures –which are likely to align with predicates anyway (see for example Figure 5). Therefore, the analysis of exact temporal alignment needs to be explored in more detail to test McNeill's (1992) suggestions more directly. Furthermore, the rest of the gesture scale should also be examined since this study has solely focused on viewpoint gestures. Finally, other properties of events allowing for either more CVPT or more OVPT representations (as proposed and discussed by Parrill, 2009) will also have to be explored in more detail.

To conclude, this study has shed more light on the co-ordination between speech and gestures at the discourse level. Specifically, we have provided the first quantitative evidence that both modalities work together to indicate the accessibility of referents through a variation of linguistic forms, on the one hand, and through the variation of viewpoint in representational gestures, on the other hand. Thus, gestural viewpoint can be seen as a cohesive device in narrative discourse, which makes gesture in general and gestural viewpoint in particular a discursive phenomenon. Our study shows how important it is to go beyond looking at the co-ordination between speech and gesture on the word or sentence level, and look at the complex interplay of the two modalities in connected discourse. It is only by considering the functions of gestures in connected discourse that we will further our understanding of the integrated nature of speech and gesture.

REFERENCES

Ahrenholz, Bernt (2005). Reference to persons and objects in the function of subject in learner varieties. In H. Hendriks (Ed.). *The Structure of Learner Varieties* (pp. 19-64). Berlin/New York: Mouton de Gruyter.

Ahrenholz, Bernt (2007). *Verweise mit Demonstrativa im gesprochenen Deutsch. Grammatik, Zweitspracherwerb und Deutsch als Fremdsprache*. Berlin & New York: de Gruyter.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics, 24,* 65-87.

Ariel, M. (1990). *Accessing noun-phrase Antecedents.* London: Routledge.

Ariel, M. (1994). Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics, 30,* 3-42.

Ariel, M. (1996). Referring expressions and the +/- coreference distinction. In T. Fretheim & J. K. Gundel (Eds.), *Reference and referent accessibility* (pp. 13-35). Amsterdam: John Benjamins.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schliperoord & W. Spooren (Eds.), *Text representation* (pp. 29-87). Amsterdam: John Benjamins.

Baayen, R. H. (2008) *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chafe, W. (1994). *Discourse, consciousness, and time.* Chicago: Chicago University Press.

Clancy, P. (1980). Referential choice in English and Japanese. In W. Chafe (Ed.), *The pear stories: Cognitive, cutural, and linguistic aspects of narrative production* (pp.127- 202). Norwood, NJ: Ablex.

Clark, H. (1996). *Using Language.* Cambridge: Cambridge University Press.

Cook, S. W., Yip, T., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and cognitive processes, 2012, 27,* 594-610.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1-42). Amsterdam: John Benjamins.

Givón, T. (1984). Universals of discourse structure and second language acquisition. In W. E. Rutherford (Ed.), *Language universals and second language acquisition* (pp. 109- 136). Amsterdam: John Benjamins.

Givón, T. (1985). *Functionalism and grammar.* Amsterdam: John Benjamins.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish.* Doctoral dissertation, Lund University.

Gullberg, M. (2003). Gestures, referents, and anaphoric linkage in learner varieties. In C. Dimroth & M. Starren (Eds.), *Information structure, linguistic structure and the dynamics of language acquisition* (pp.311-328). Amsterdam: John Benjamins.

Gullberg, M. (2006). Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning, 56,* 155-196.

Gundel, J. K., Hedberg, N. & Zacharaski, R. (1993). Cognitive status and the form of referring expressions. *Language, 692,* 274-307.

Hickmann, M. & Hendriks, H. (1999). Cohesion and anaphora in children's narratives: a comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language, 26,* 419-452.

Hendriks (2003). Using nouns for reference maintenance: A seeming contradiction in L2 discourse. In A.G. Ramat (Ed.), *Typology and second language acquisition* (pp.291- 326). Berlin: Mouton de Gruyter.

Huang, Y. (2000) *Anaphora. A cross-linguistic study.* New York: Oxford University Press.

Jancovic, M., Devoe, S., & Wiener, M. (1975). Age-related changes in hand and arm movements as nonverbal communication: Some conceptualizations and an empirical exploration. *Child Development, 46(4)*, 922-928.

Kendon (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kita (2000). How representational gestures help speaking. In D. McNeill (Ed.), Language and Gesture. (pp.162-185). Cambridge: Cambridge University Press.

Kita, S. & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48,* 16-32.

Lambrecht, K. (1994). *Information structure and sentence form.* Cambridge: Cambridge University Press.

Levy, E. T. & Fowler, C.A. (2000). The role of gestures and other graded language forms in the grounding of reference in perception. In D. McNeill (Ed.), *Language and gesture* (pp. 215 234). Cambridge: Cambridge University Press.

Levy, E. T., & McNeill, D. (1992). Speech, gesture, and discourse. *Discourse Processes 15,* 277-301.

Marslen-Wilson, W., Levy, E. T., & Komisarjevsky Tyler, L. (1982). Producing interpretable discourse: The establishment and maintenance of reference. In R. J. Jarvella & W. Klein (Eds.), *Language, place, and action. Studies in deixis and related topics* (pp. 271-295). Chicester: John Wiley.

McNeill, D. (1992). *Hand and Mind: what gestures reveal about thought.* Chicago: Chicago University Press.

McNeill, D. & Levy, E. (1993). Cohesion and gesture. *Discourse Processes, 16,* 363-86.

Parrill, F. (2009). Dual viewpoint gestures. *Gesture 9,* 271-289.

Parrill, F. (2010). Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes, 25(5),* 650-668.

Parrill, F. (2012). Interactions between discourse status and viewpoint in co-speech gesture. In B. Dancygier & E. Sweetser (Eds.), *Viewpoint in Language: A Multimodal Perspective.* Cambridge: Cambridge University Press.

Perniss, P. (2007). *Space and Iconicity in German Sign Language (DGS).* MPI Series in Psycholinguistics 45, Radboud University Nijmegen.

Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture.* MPI Series in Psycholinguistics 39, Radboud University Nijmegen.

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category - ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation.* (LREC 2008).

So, W.-C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science, 33,* 115-125.

Yoshioka, K. (2005*) Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.* Doctoral Dissertation, University of Groningen.

APPENDIX

Film 1 "Die Sendung mit der Maus" (1972): The pancake event

(Description adapted from Perniss, 2007)

1. The mouse is standing at a kitchen stove with its back to the viewer, its arms and shoulders moving
2. The mouse turns away from the stove to the left and the viewer sees that it is holding a pan with a pancake in it
3. The mouse flips the pancake, but instead of landing in the pan, the pancake arcs through the air and lands on the floor
4. The mouse moves forward and bends down to pick the pancake up off the floor
5. The mouse turns back to the stove to prepare a new pancake (this sequence repeats two more times)
6. The fourth time the mouse turns away from the stove to try to flip the pancake and runs forward to catch the pancake in the location in which it has been landing
7. This time the pancake lands on the mouse's forehead
8. The mouse then calls the elephant for assistance, who enters from the screen from the left
9. The mouse motions for the elephant to stand at a distance, and they are standing across from each other (facing each other)
10. The mouse moves toward the elephant to hand it the other pan hanging on the wall
11. The mouse flips the pancake, and it arcs through the air and lands in the elephant's pan
12. The elephant flips the pancake, and it arcs back through the air and lands in the mouse's pan

13. The pancake moves back and forth between the mouse and the elephant three times, until the screen fades

Film 2: "Die Sendung mit der Maus"(1972): The ball event

(Description adapted from Perniss, 2007)

1. The elephant is lying on the ground on its stomach, facing away from the viewer, while the mouse is on the right side of the screen, facing the elephant
2. The mouse moves toward the elephant, dribbling the ball and gets the elephant's attention
3. The elephant turns to the right, toward the mouse and the gets up and stands across from the mouse, facing the mouse, in order to play ball
4. The mouse starts the game by throwing the ball to the elephant and the elephant uses its trunk to hit the ball back to the mouse
5. The ball passes back and forth between the mouse and the elephant in this manner five times
6. The mouse then kicks the ball to the elephant, the elephant hits the ball back with its trunk, kicking his legs out as it does so
7. The ball passes back and forth between them in this way two times
8. The mouse then uses its nose to hit the ball and the elephant, in turn, stretches its trunk high into the air to hit the ball
9. The ball passes back and forth between them in this way two times
10. Then suddenly, the elephant swallows the ball instead of hitting it back
11. The mouse walks over to the elephant
12. The mouse tickles the elephant's trunk, causing the elephant to sneeze
13. The ball flies out of the elephant's trunk and hits the mouse in the stomach
14. From the impact of the ball, the mouse flies backward though the air

Film 3: "Die Sendung mit der Maus"(1972): The bar event

(Description adapted from Perniss, 2007)

1. The mouse is standing in front of a gymnast's bar
2. The mouse moves toward the bar, until it is standing under it
3. The mouse jumps up to grab a hold of the bar, and starts swinging back and forth on it
4. The elephant enters the screen from the right and watches the mouse as it starts to swing around the bar and then as she dismounts from the bar
5. The mouse turns around to look at the elephant and moves away from the bar, such that the mouse and the elephant are in either side of the bar
6. The elephant also wants to swing on the bar and moves toward it
7. The elephant reaches the bar with its trunk, but the bar bands under the elephant's weight
8. The elephant lets go of the bar, swinging back onto the ground
9. Furious, the mouse moves toward the bar to try to straighten it, but is unable to do so and moves away from the bar again
10. A man dressed in a green suit and top hat enters the screen from the left and walks under the bar, straightening the bar as he does so
11. Amazed, the mouse and the elephant look at each other and then at the viewer, standing on each side of the bar

Film 4: "Die Sendung mit der Maus"(1972): The drums event

1. The mouse is sitting on an armchair reading a jurisprudential book on the right side of the screen
2. The elephant comes in from the left side of the screen with drums in his hands
3. The elephant sits down facing the viewer and starts playing the drums, still sitting on the left side of the screen
4. Appalled, the mouse jumps up in her chair, the book falls out of its hands
5. The mouse is disturbed by the noise and looks angrily to the elephant

6. Furious, the mouse gets up from her chair and walks toward the elephant
7. The mouse takes the beaters away from the elephant and throws them away toward the right side of the screen (the beaters disappear)
8. The mouse walks back to her chair, sits down, picks up its book and continues reading
9. The elephant starts beating the drum with his hand and trunk
10. Again, the mouse puts away the book, looks up, is angry, gets up and walks to the elephant
11. This time, the mouse takes away the drum and throws it away toward the right side of the screen (the drum disappears)
12. The mouse goes back to the chair, sits down, picks up the book and continues reading
13. The elephant looks around, seemingly thinking about what to do next
14. The elephant then starts beating with his hands on his belly
15. The mouse looks up, again being annoyed by the noise
16. The mouse takes its ears, pulls them up and knots them into a tie, while the elephant looks at the mouse and to the viewer in turns
17. The mouse then picks up the book again and starts reading
18. The elephant continues to beat on his belly

Film 5: "Die Sendung mit der Maus" (1972): The hippo event

1. The elephant and a duck are walking facing the left side of the screen
2. The duck is behind the elephant, they have the same size
3. The duck jumps up five times while walking and screams
4. The duck then jumps onto the elephant's back
5. The elephant stops walking and looks up to the duck
6. The duck screams twice
7. The elephant moves his body to make the duck go away while the duck is shaky but rests on his back
8. The elephant takes a deep breath and then continues walking with the duck on his back

9. The elephant stops walking
10. The elephant and the duck lean forward
11. They seemingly see something
12. The duck screams twice
13. Then a hippo slowly walks into the picture from the left side of the screen
14. The hippo walks by the elephant and the duck
15. The duck jumps from the elephant's back onto the back of the hippo
16. The elephant follows: it jumps on the hippo's back
17. The duck then jumps on the elephant's back again
18. Now, the three animals form a pyramid, the hippo carries the elephant who in turn carries the duck
19. The hippo continues walking
20. Then the hippo stops
21. The elephant and the duck look down to see what is going on
22. The hippo turns toward the viewer
23. The duck screams twice
24. Then the hippo turns back facing to the right side of the screen and continues walking

Film 6: "City Lights" (1931) (Charlie Chaplin): The date

(Description adapted from So et al. 2009)

1. A cat wags its tail while sitting on a windowsill
2. Charlie Chaplin gets out of a car while a woman sits idly by
3. Charlie Chaplin is holding a basket and leads the woman down a corridor
4. Charlie Chaplin doffs his hat towards the woman
5. Charlie Chaplin gives woman a basket
6. Charlie Chaplin grabs woman's hand
7. Charlie Chaplin kisses woman's hand
8. The woman walks up the steps while Charlie Chaplin watches her

9. The woman opens the door to her apartment
10. The cat knocks a pot off the windowsill
11. The falling pot strikes Charlie Chaplin on his head and he falls down

Film 7: "City Lights" (1931) (Charlie Chaplin): The good deed

(Description adapted from So et al. 2009)

1. A man holding a suitcase descends a flight of steps
2. The man opens the suitcase
3. The man throws noose around his neck and then opens it
4. The man continues to hold noose around neck while he watches Charlie Chaplin with a cane descend the steps
5. The man drops a rock on the foot of Charlie Chaplin
6. Charlie Chaplin removes noose from the man
7. The man throws noose back around his neck as well as around Charlie Chaplin
8. The man falls from under noose while Charlie Chaplin looks off in the other direction
9. The man throws rock into a nearby body of water
10. Charlie Chaplin pulls the man into water with him